Enterprise Vault Best Practices

Implementing SharePoint Archiving

This document contains information on best practices when implementing Enterprise Vault for SharePoint

If you have any feedback or questions about this document please email them to **IIG-TFE@symantec.com** stating the document title.

This document applies to the following version(s) of Enterprise Vault: 9 & 10

This document is provided for informational purposes only. All warranties relating to the information in this document, either express or implied, are disclaimed to the maximum extent allowed by law. The information in this document is subject to change without notice. Copyright © 2012 Symantec Corporation. All rights reserved. Symantec, the Symantec logo and Enterprise Vault are trademarks or registered trademarks of Symantec Corporation or its affiliates in the U.S. and other countries. Other names may be trademarks of their respective owners

Document Control

Contributors

Who	Contribution
David Scott	Contributor
Evan Barrett	Contributor & Author

Revision History

Version	Date	Changes
1.0	November 2011	Initial release
2.0	February 2012	Policy best practice updates
3.0	July 2012	Updates for Enterprise Vault 10.0.2

Related Documents

Version	Date	Title

Table of Contents

Introduction

This document will cover best practices for implementing successful Enterprise Vault SharePoint Archiving and will include hint and tips. This document is not an introduction to SharePoint archiving and assumes that the reader will have prior knowledge of Enterprise Vault and SharePoint.

Enterprise Vault Server Configuration

Operating System and Hardware

Starting with Enterprise Vault 10, only Windows 2008 R2 and later will be supported on the Enterprise Vault server itself. The hardware requirements for Enterprise Vault 10 have also changed when compared to previous versions. For absolute hardware minimums, there must be at least four CPU cores and 8GB RAM in the Enterprise Vault server.

For best indexing performance, it is recommended to increase the RAM to 16GB or higher. A benefit of increased memory is that an Enterprise Vault server can keep more index data in memory which can reduce the amount of disk I/O during searches.

It is highly recommended that Index Volumes be placed on high speed disk such as a storage area network (SAN) or direct attached storage (DAS). This will reduce the amount of time required when searching archived file content due to the fact that Enterprise Vault will be able to obtain data from disk more rapidly.

The temp location for the Vault Service Account (VSA) should also be placed on high speed disk. This will increase the archiving throughput. The size of the disk volume should be large enough to handle the largest files in the environment that will be archived. Any file larger than 50MB in size will be temporarily transferred to this area for fingerprinting (a file less than 50MB in size will be fingerprinted in place on the file server).Thus, if the environment has 5 GB documents that will be archived, more than 5GB of disk space will be needed for the VSA temp location.

By default the VSA temp area will be location on the operating system drive (normally C:). As well as ensuring this is on high speed disk, it is recommended to move the VSA temp area away from the operating system drive.

Domain Name Service (DNS)

It is necessary to ensure that all Enterprise Vault and target SharePoint servers are correctly registered in DNS. A properly configured DNS environment will reduce the amount of errors when setting up SharePoint archiving.

Indexing

Enterprise Vault offers the ability to defer indexing with SharePoint, but this is not recommended for a couple of reasons. Some SharePoint tools, such as EVSPShortcutManager, require indexes in order to function. If search (using Browser Search or Discovery Accelerator) and/or the use of Archive Explorer are required, indexing must not be deferred as these features access SharePoint indexes. In these cases, only minimal indexing is required (indexing level set to Brief).

The index storage requirements for SharePoint are less when compared to other content (such as Exchange) that can be archived by Enterprise Vault. For SharePoint archiving, a Brief level for indexing will typically require around 2% of the original file size and a Full level will typically require around 6% of the original file size.

What is installed

- EV for SharePoint is installed on every front-end server in a SharePoint farm. Running the EV install for SharePoint will install the following: HTTP Module
- Web service extensions
    - This is used to extend the SharePoint web services with some additional methods that provide functionality that isn't there in the core web services (i.e. to support the remote archiving task). This is all done by the installer – no post install configuration is necessary.
- EV for SharePoint Search Web Part
    - A custom Web Part for Enterprise Vault Archive Search is added to the Web Part gallery during installation of the Enterprise Vault SharePoint components.
    - To enable users to search Enterprise Vault archives, you need to add the Archive Search Web Part to SharePoint page
- EV Version History
    - The original SharePoint version history is replaced with the EV for SharePoint Version history (allowing view of archived versions of the document in addition to versions still in SharePoint).
- EV Admin Service
    - Installs the EV admin service to provide required support for the Index Client (i.e. supports the Archive Search web app)

NOTE: To check that EV for SharePoint is properly installed, refer to the following tech note***:*** **http://www.symantec.com/docs/HOWTO60773**

Permissions on SharePoint

Document Permissions

Enterprise Vault will keep track of permissions at the folder level and not for individual objects such as files when archiving from SharePoint targets. Files whose permissions are different than that of the parent folder will have the option of either being archived with the parent folder permissions or not archived at all. This option can be set in the Permissions tab of the SharePoint policy.

To determine if files have different permission than that of the parent folder, run the SharePoint Task against the desired SharePoint target in Report Mode (with verbose logging enabled). The report, which will be stored in the \Enterprise Vault\Reports\SharePoint Archiving, will identify any files that have explicit permissions. If these files are to be archived, the seamless shortcut on the SharePoint will still have its original permissions, but the file will have the parent folder's permissions within the archive on Enterprise Vault.

SharePoint Permissions

By default, the Enterprise Vault SharePoint task uses the Vault Service Account (VSA) to access SharePoint. However, this can be changed by modifying the SharePoint task in the Vault Administration Console (VAC). The account that the Enterprise Vault uses to access SharePoint **must have site collection administrator** permissions on the target SharePoint site collections. To configure the account for site collections administrator, use the SharePoint Central Administration web page. Navigate to Application Management-> User Policy.

For 2010:

1. Open the "SharePoint Central Admin"
2. Click on "Manage Web Applications" under "**Application Management"**
3. Click on the Web application that you wish to change
4. Click on the User Policy button in the ribbon.
5. Add the Vault Service Account with Full Control Permissions

For 2007:

1. Go to the "Policy for Web Application" setting in SharePoint Central Administration.
2. Add the Enterprise Vault SharePoint task account with Full Control permissions

Now the Enterprise Vault SharePoint task has full access to all the site collections in the Web application

Defining a SharePoint Target with Enterprise Vault

Overview

Enterprise Vault for SharePoint can set policies at the web application, site collection or sub-site level. The target allows you to define different policies and retention. It is important to factor this into the decision on how granular targets need to be. If all content can share the same retention category, such as in a storage optimization scenario, the Enterprise Vault administrator can simply set the target at the Web application level and have all sub-sites inherit the same settings. If a sub-site needs specific policies the administrator can define the archiving policy for the sub-site separately.

What Is a Web Application in SharePoint?

Before you can create a site or a site collection, you must first create a Web application. A Web application is comprised of an Internet Information Services (IIS) site with a unique application pool. When you create a new Web application, you also create a new database and define the authentication method used to connect to the database. The following screenshot illustrates setting up a SharePoint Web Application target in Enterprise Vault. If the SharePoint Web Application does not use port 80, the port number will need to be added to the URL. Figure 1 illustrates how to define a SharePoint target.

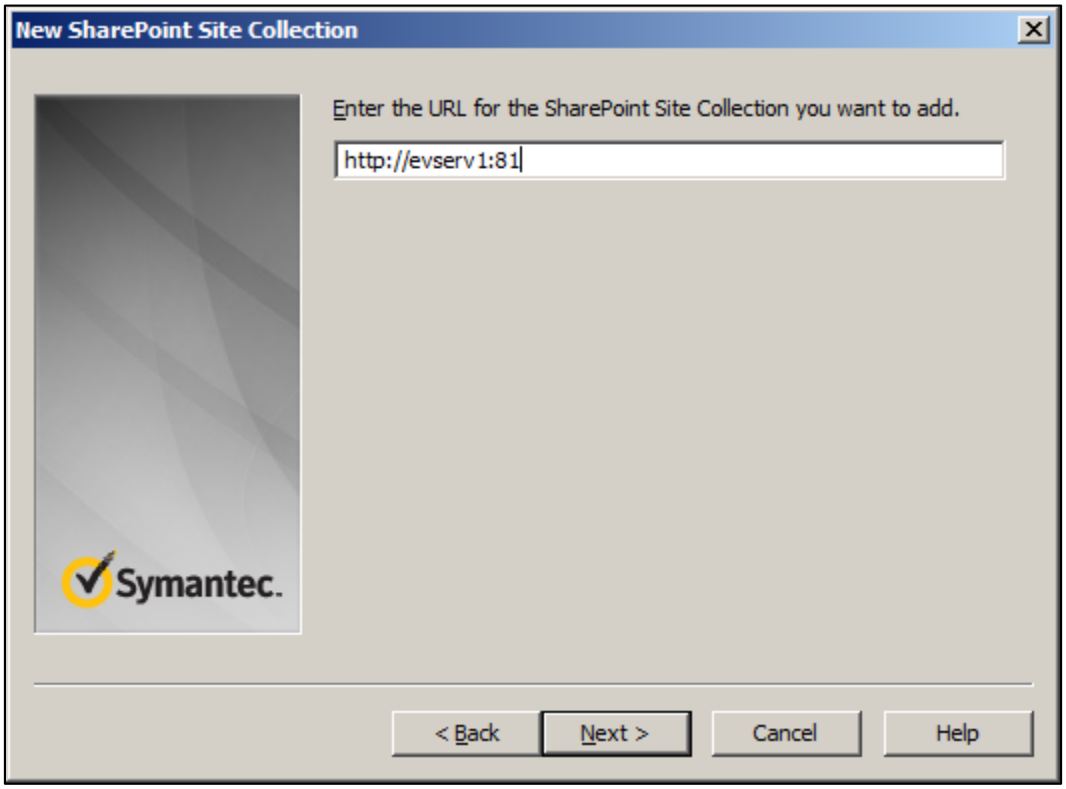

Figure 1 - Defining a Web Application target

What Is a Site Collection?

Site collections in SharePoint consist of a set of web sites on a virtual SharePoint server that have the same owner and share administration settings. Each site collection contains a top-level Web site and can contain one or more sub-sites. The only role that has permission over the entire collection of sites is the Site Collection Administrator. Figure 2 illustrates how to configure a site collection target within Enterprise Vault.

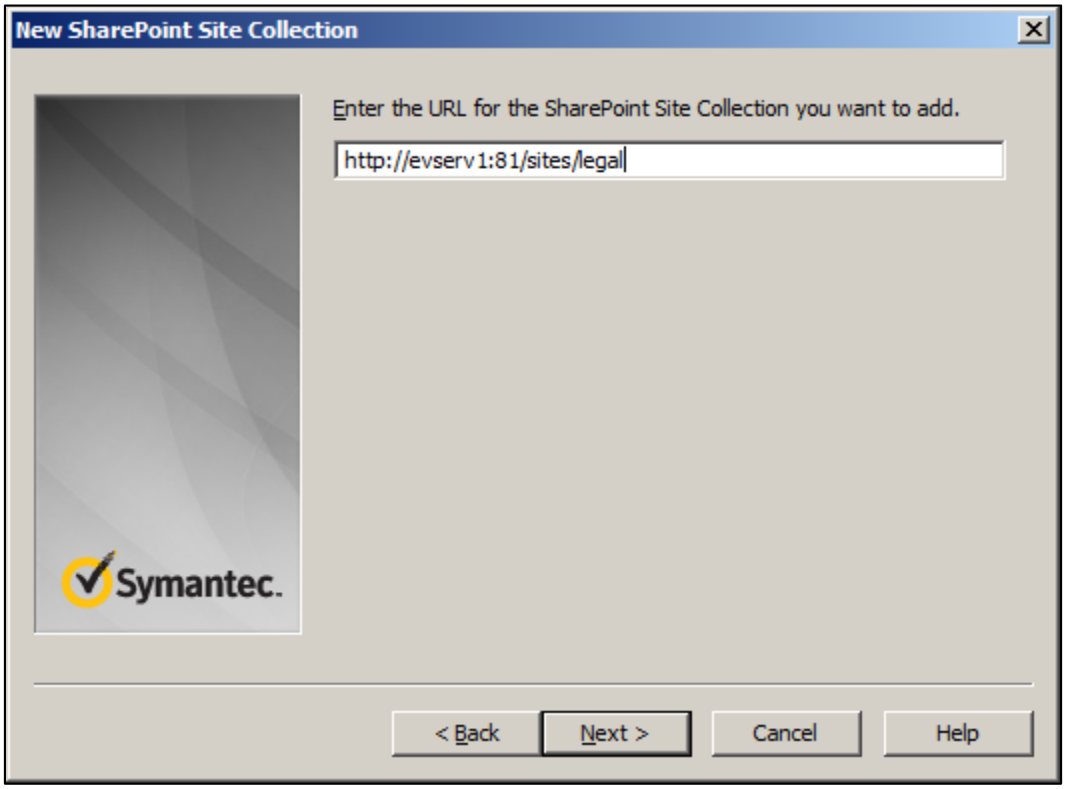

Figure 2 - Defining a Site Collection target

What Is a Sub-site?

A sub-site is a complete web site stored in a named subdirectory of the top-level Web site. Each sub-site can have administration, authoring and browsing permissions that are independent from the top-level Web site and other sub-sites. A sub-site can also have sub-sites of its own. Because every site below the top-level site is actually a sub-site, each sub-site is generally called simply a site. To open a sub-site, you supply the URL of the server and any sub-sites—for example, http://evserv1:81/sites/legal/Litigation — without specifying a page name. Figure 3 illustrates how to define a sub-site target within Enterprise Vault.

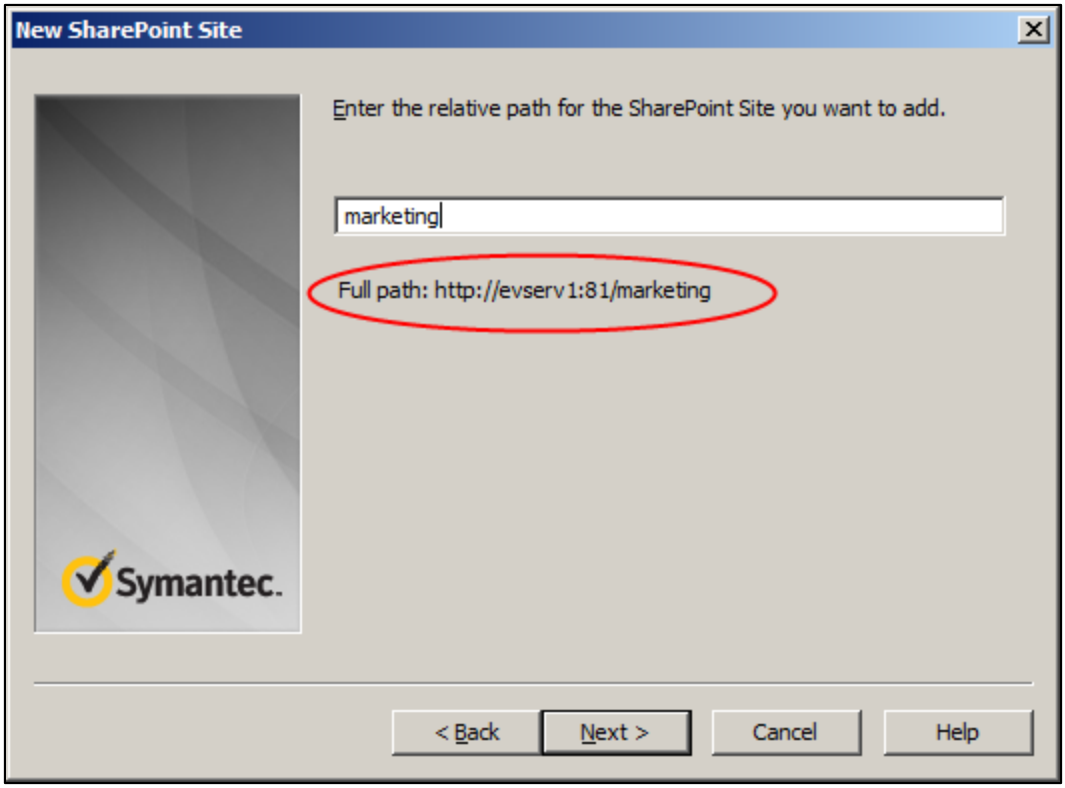

Figure 3 - Defining a Sub-site target

When Are Granular Targets Needed?

Target a site collection or sub-site if one of the following scenarios are required:

- A specific retention category is required
- A different archiving policy is required for the target
- A different SharePoint archiving task is needed (ex. You may wish to have a more aggressive archiving policy)
- A different Vault Store is need (available only at the web application target level)

Figure 4 illustrates how options can be set for a target.

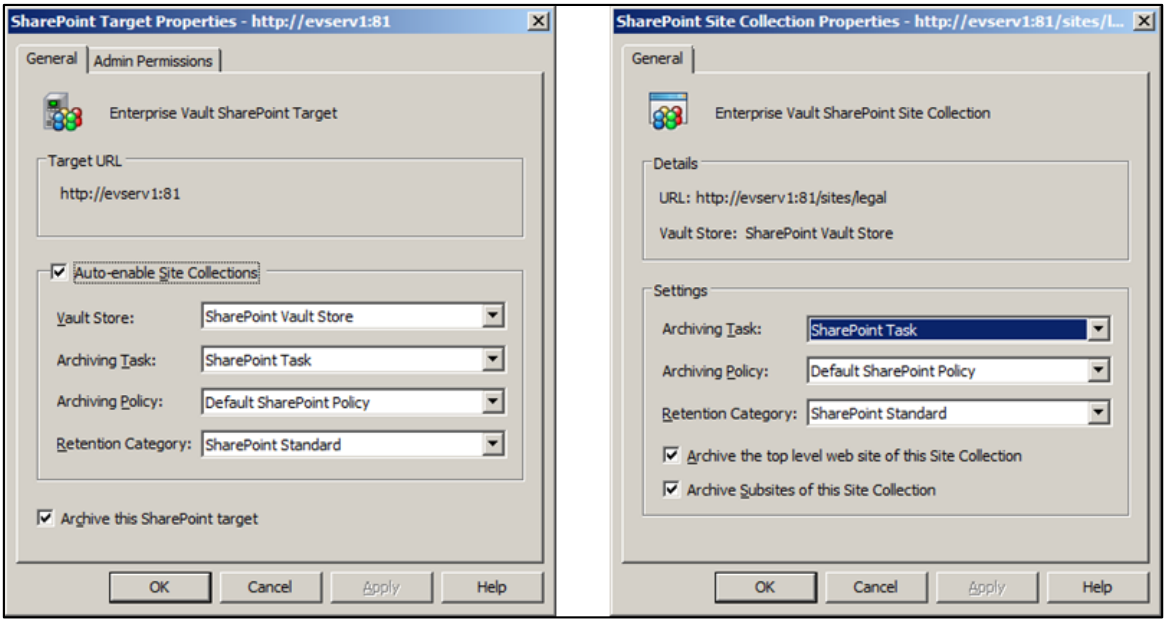

Figure 4 - SharePoint Target options

Creating SharePoint Archiving Policies

Know the Content before Archiving

Before implementing SharePoint archiving, it is a good idea to know the nature of the content that resides in the target SharePoint sites. Implementing a policy that will essentially "archive everything" (e.g.: using *.* as the file filter in the policy) is not recommended as files may be archived that are not great candidates for archival. A utility from Symantec, SharePoint Analyzer (available for both SharePoint 2007 and 2010), is available that will scan document libraries within SharePoint sites and collect metadata on all discovered files. This metadata includes file type, file size, and date (last modified). The output from this utility can be used to design proper SharePoint archiving policies.

Efficient Policies Equate to Better Archiving Performance

When designing SharePoint archiving policies, it is important to note that the more rules that there are in in a policy, the longer it will take to archive SharePoint content. Therefore, it is important to design efficient archiving policies with the minimal amount of rules necessary to archive content. If trying to archive different file types, such as Microsoft Office documents and PDF files, create one rule that incorporates all the file types (*.doc,*.xls,*.ppt,*.pdf, etc.) instead of creating two separate rules.

Archiving "Everything" Policies

There are many files that could reside in a SharePoint environment that should not be archived such as files that are frequently modified. Implementing an "archive everything" policy may result in numerous retrieval requests from the Enterprise Vault server.

There are times when using an archive "everything" rule makes sense. Use cases for this include:

- Storage Optimization – while it is possible that you would want to blindly archive everything and replace with a shortcut, it is perhaps more likely that this would be limited to only archive content older than a specific date
- A copy of files destined for preservations purposes (ex. Archive a sub-site prior to deleting it from SharePoint)
- If there are other "do not archive" rules higher in the hierarchy in an archiving policy
- When archiving all files is required for eDiscovery and compliance

Deleting Files

Enterprise Vault for SharePoint has the ability to delete unneeded files during archiving that can take up a lot of space but offer little to no compliance or discovery benefit. These files typically include temporary files created by an application. SharePoint administrators may also find it necessary to delete other file types in order to comply with company or government regulations. An example of this may be to delete all MP3 files as there may be a policy to not store these types of files on company hardware resources.

Small Files

Unless there is a need to archive a small file for compliance purposes, there is no benefit to archiving small files. A placeholder will take approximately 6 bytes on the SharePoint server. Any file archived by Enterprise Vault will consume around 10KB of space on the Enterprise Vault server itself. This space is divided between SQL, index, and saveset. Generally, files that are less than ~11KB will not save any disk space in the environment overall unless there are many duplicates of file. An exclusion rule for small files should be placed in the top of the archiving policy. Figure 5 and Figure 6 (EV 10.0.2 and later) illustrate how to define a rule that would not archive smaller files.

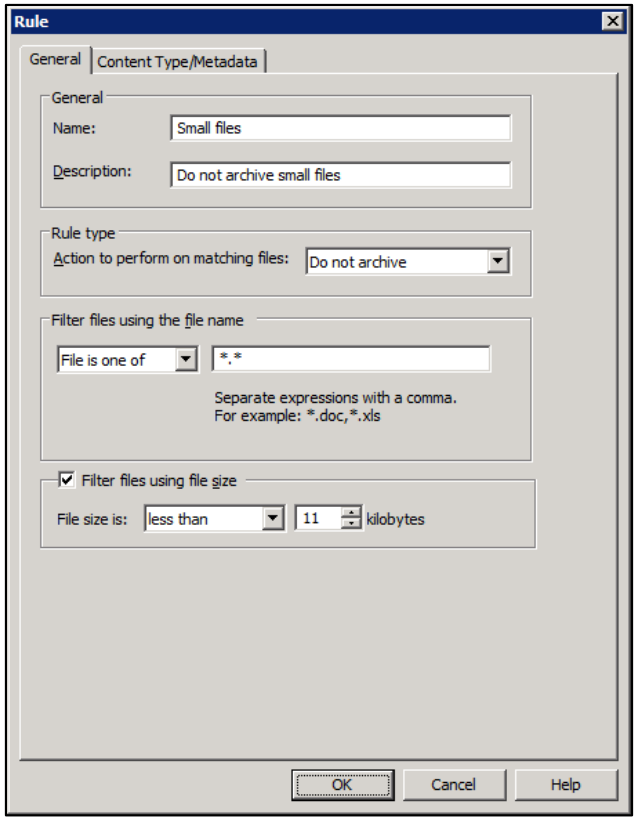

Figure 5 - Rule for not archiving small files

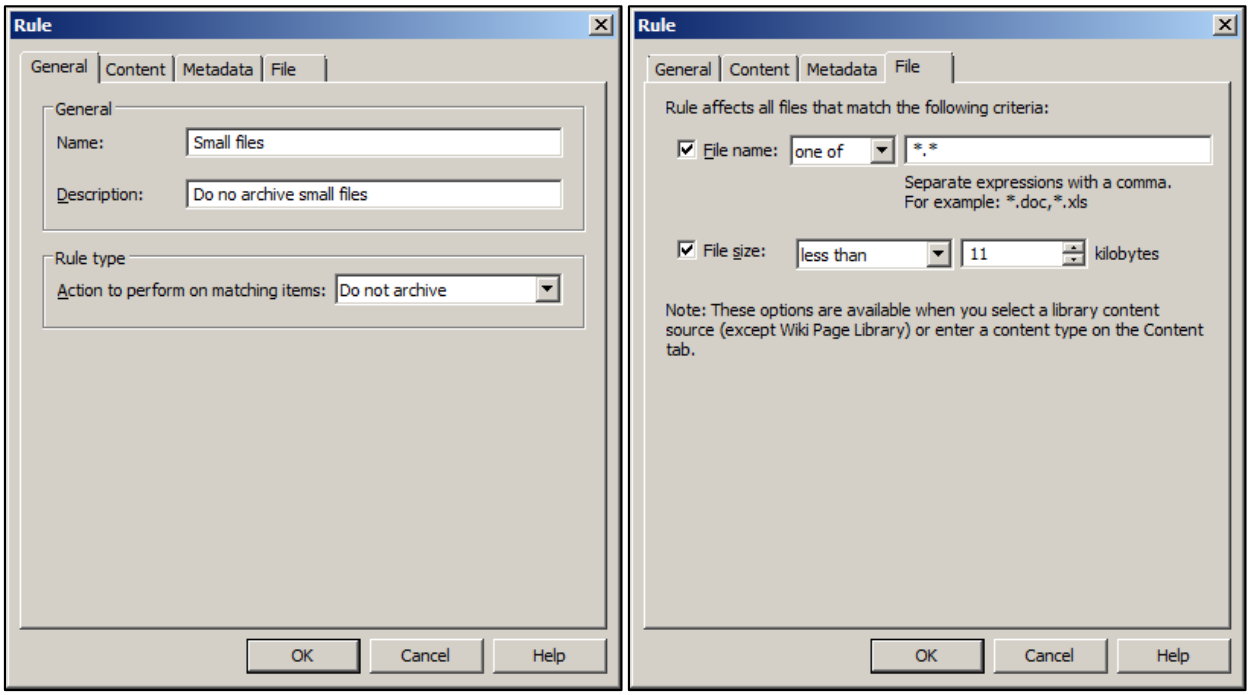

Figure 6 - Rule for not archiving small files (EV 10.0.2 and later)

Large Files

Enterprise Vault has the ability to archive documents of any size. However, it should be noted that large documents will take time when being retrieved from Enterprise Vault. End users should be educated on this fact in order to set expectations. If space reduction on SharePoint servers is the main goal of implementing Enterprise Vault for SharePoint, it is recommended that only larger documents initially be archived. This will show the greatest space savings on the SharePoint server. Once these larger documents have been archived, the archiving rules can be changed to allow smaller documents to be archived. Figure 7 and Figure 8 (EV 10.0.2 and later) illustrate how to only archive documents that are larger than 5MB.

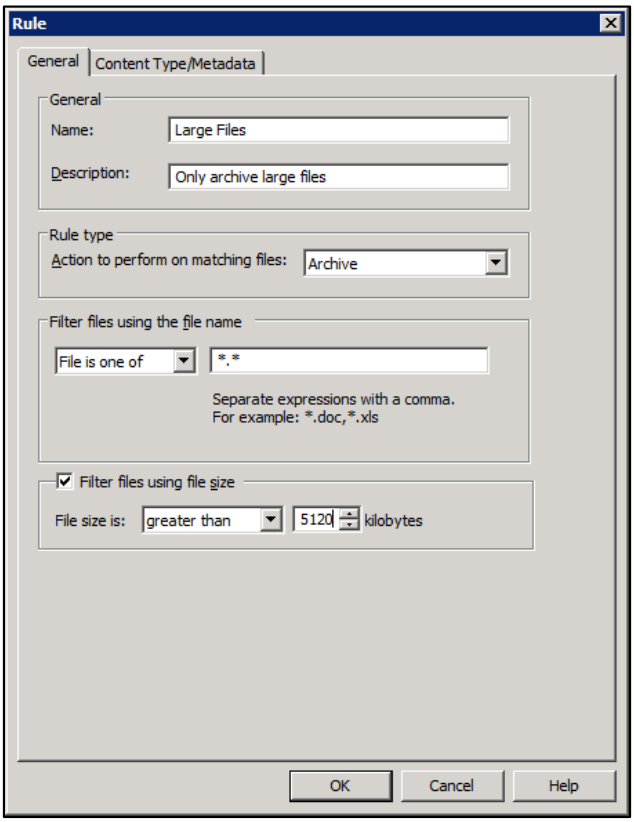

Figure 7 - Rule for archiving large files

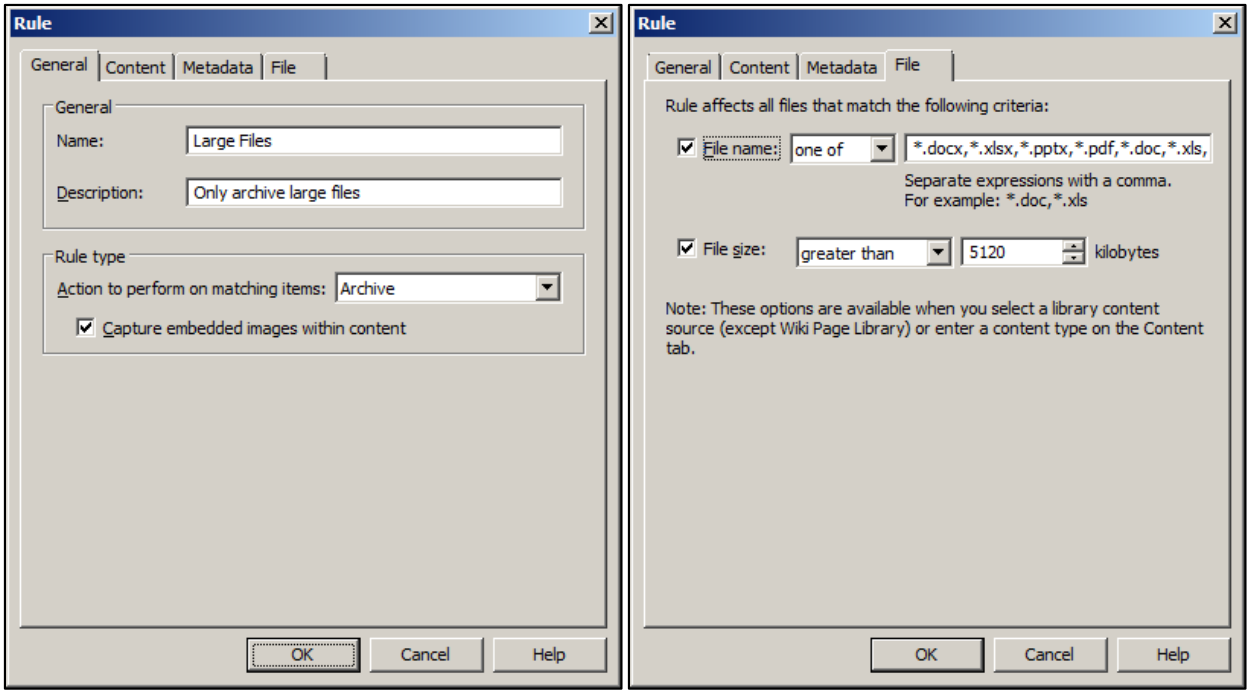

Figure 8 - Rule for archiving large files (EV 10.0.2 and later)

Dates of Documents

Enterprise Vault for SharePoint allows archiving based on last modified date, creation date, and any custom date field. Last accessed time is not available as this metadata type is not offered in SharePoint. For best practices, it is recommended that a non-aggressive policy be used initially by only archiving older documents (such files that are two or three years old). Once these older files have been archiving, the age of the files can be reduced (e.g. two years to one year, one year to six months, etc.). This type of policy will provide the least impact to end users and SharePoint servers and will also allow the administrator to tune the policy based on the actual space gained. Figure 9 illustrates how to define a rule that will only archive documents older than one year.

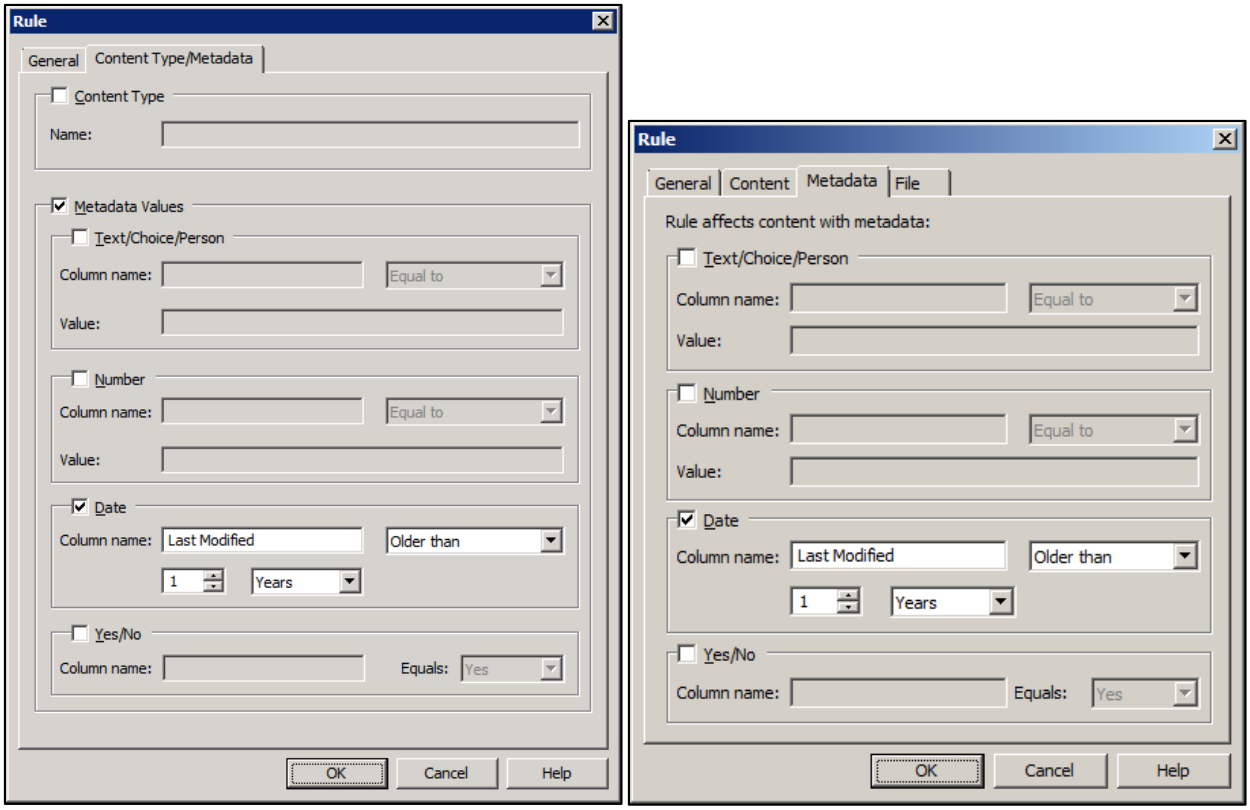

Figure 9 - Archiving based on modification date (EV 10.0.2 and later on right)

Best Practices for Shortcuts

Seamless shortcuts on the SharePoint server are six bytes in length and redirect requests to the content which now resides within the archive. Archiving does not create a shortcut automatically, instead content is archived and remains in SharePoint by default. To remove the original content and/or replace it with a shortcut, the administrator is required to define the shortcut policy.

You have three options for the shortcut policy:

1. Remove Content that has been archived and replace it with a shortcut (End-user experience is maintained via shortcuts but actual item now resides in Enterprise Vault)
2. Remove Content and don't replace with a shortcut (Content is completely gone from SharePoint and only exists in Enterprise Vault. In this scenario the end user can only access archived content via EV Web Search, the EV for SharePoint Search Web part and Archive Explorer).
3. Don't remove items from SharePoint after they have been archived (SharePoint content is essentially duplicated in the archive).

If you wish to replace an item with a shortcut as soon as it has been archived, a zero-day shortcut policy can be used. Shortcuts can also be deleted from the SharePoint server after a specified time. The

archived content will still be available on the Enterprise Vault server and will be retained for the duration of the retention category that was assigned to the item at the time of archive. The EV integrated web part, Browser Search, or Archive Explorer can be used to retrieve the item. Deleting older shortcuts from the SharePoint server will remove stale content from the SharePoint index and will accelerate the backup of SharePoint.

Starting with Enterprise Vault 8.0SP3, shortcuts are seamless. The URL for the archived item does not change. This means that references to the item before archiving are still viable. Editing of an archived item is still allowed. If the item should be modified and re-saved back to SharePoint, the shortcut will be replaced with the new version of the document and would be archived again once it qualifies (based on the archiving policy). Archived content is fully indexed by SharePoint as long as the shortcut is present. Explorer view is also preserved with proper icons being displayed.

Figure 10 illustrates how to define a shortcut policy.

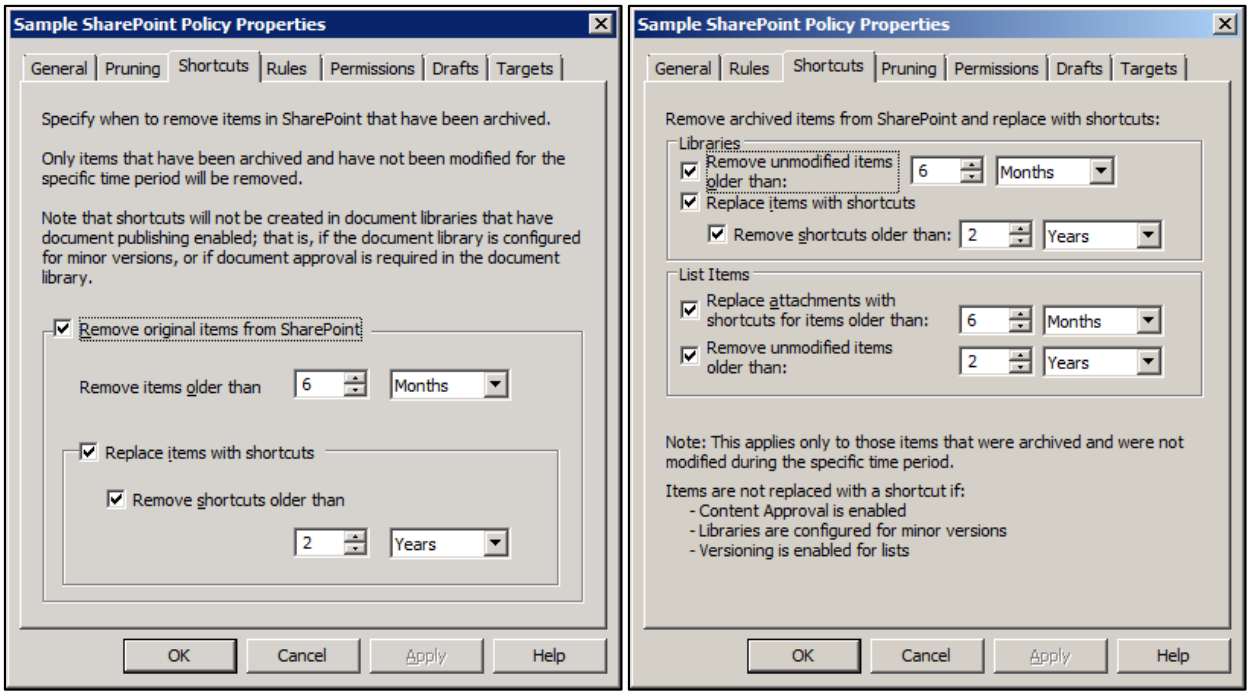

Figure 10 - Shortcut options (EV 10.0.2 and later on right)

If at a later time you decide that you wish to restore the content back into SharePoint (replacing the shortcuts) you can use a command line tool that was introduced in EV9.0.2.

Note that EV10.0.2 and later offer the ability to also remove list attachments and optionally replace them with shortcuts.

Manual (End-user Controlled) Archiving

Enterprise Vault for SharePoint can archive based on metadata. A SharePoint administrator can create an "Archive Now" column in SharePoint document libraries. This column should be a "Yes/No" type and if desired, in-place editing can be turned on to allow modification by the end user without having to first view the properties. The Enterprise Vault administrator can then create a metadata rule in the SharePoint archiving policy to act on the "Archive Now" field. The following screenshot illustrates how to set this up on both SharePoint and Enterprise Vault.

SharePoint "Create Column" UI (Figure 11):

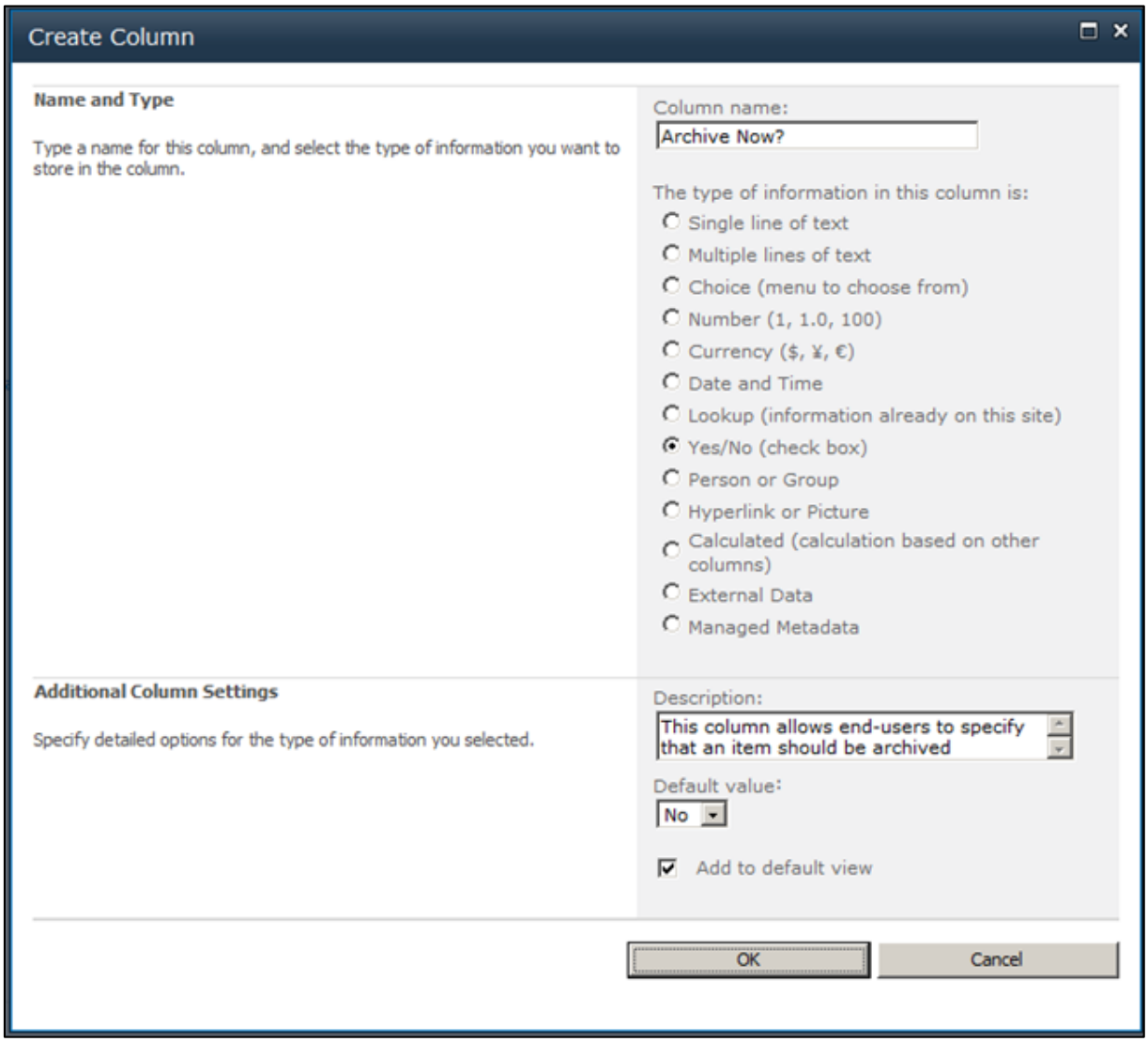

Figure 11 - Creating metadata property in SharePoint

EV for SharePoint Rule Dialog (Figure 12):

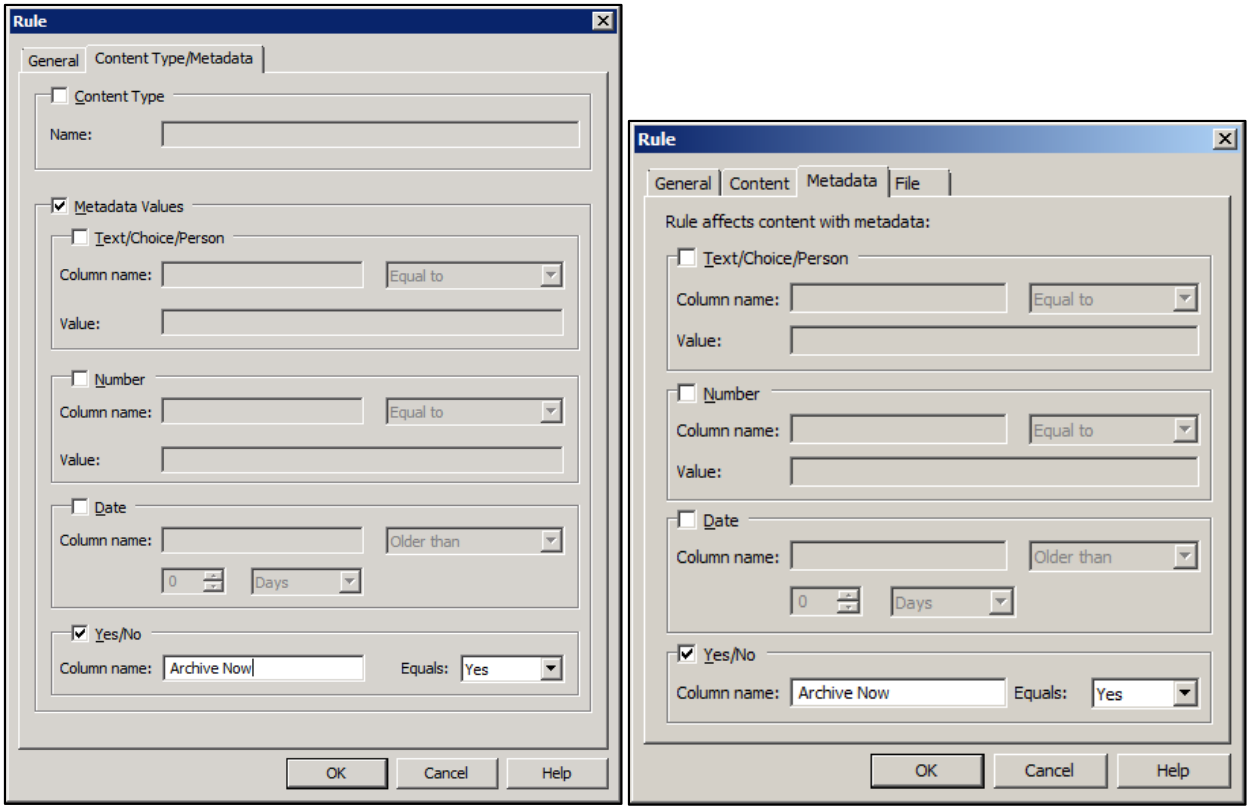

Figure 12 - Defining metadata rule for manual archiving (EV 10.0.2 and later on right)

In SharePoint, the end-user would simply click on the "Archive Now" checkbox to enable archiving as shown in Figure 13.

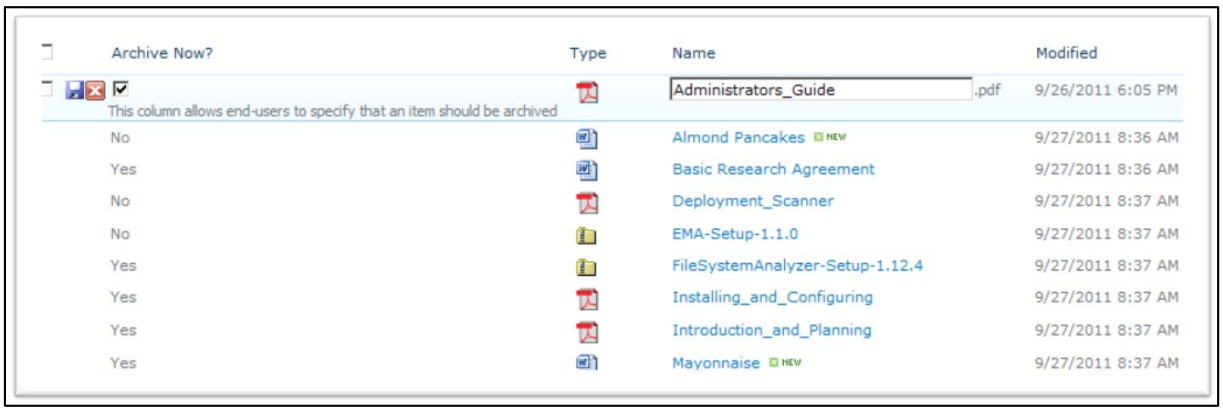

Figure 13 - Setting the Archive flag on a document in SharePoint

Archiving "Conversational Content"

SharePoint includes Blogs and Discussions which leverage posts and replies to make up a conversation. EV captures posts and replies for a given topic as a single entity in the archive to preserve the full context (i.e. seeing a post or reply in isolation does not give the full context of the conversation).

EV will bundle up all elements of the conversation during an archiving run. If you use WORM storage, this may mean you will end up with multiple versions of a document if the conversation continues after the initial archiving run. To mitigate this it is a good idea to create a specific rule for BLOGS & Discussions that includes metadata rule to only archive if "Modified Date" > 3 days (or a similar value that will give a better chance of the conversation being completed prior to archiving).

When non-WORM storage is used, this is not an issue as the Blog and WIKI will be replaced in the archive if the conversation continues.

Controlling SharePoint Version History

Within SharePoint, versioning captures every change made to a document or item. This modified file is considered a new version. Enabling versioning is one of the primary drivers of SharePoint growth. Enterprise Vault offers the ability to only keep the latest version on SharePoint while moving older versions to Enterprise Vault. End-users will still be able to access older versions on SharePoint or on the Enterprise Vault server itself using Browser Search or Archive Explorer. The following screenshots illustrate how versions of a document appear in SharePoint before and after archiving. Before pruning (Figure 14):

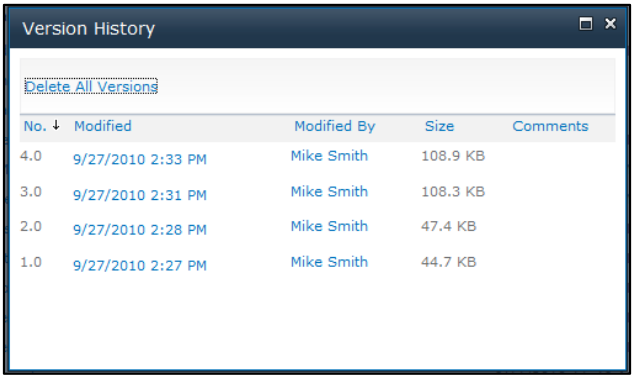

Figure 14 - Viewing version history in SharePoint before pruning

After pruning with Enterprise Vault (note how only one version resides on SharePoint and the rest are stored in the archive) as shown in Figure 15:

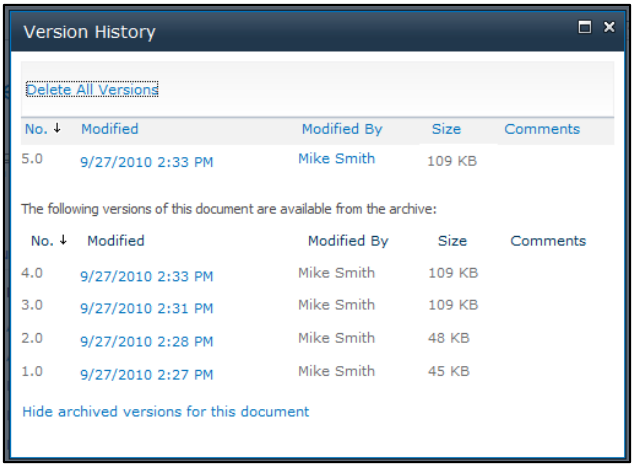

Figure 15 - Viewing version history in SharePoint after pruning

Controlling the number of versions of a document stored on SharePoint is defined in the policy as show in Figure 16.

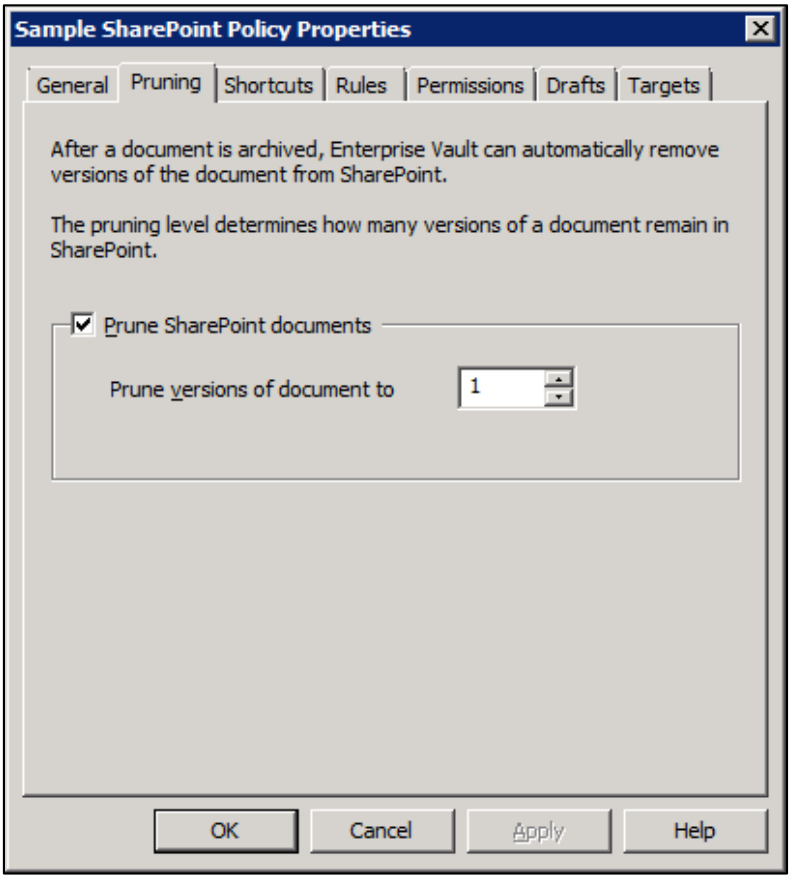

Figure 16 - Defining the pruning policy in Enterprise Vault

NOTE: EV does not prune versions until the archived copy has been backed up. So versions will be pruned on the next run after the archived copy has been backed up. This means it will always take a minimum of two runs – even if the "remove safety copies" flag is set to "immediately".

So What Does a Good Rule Hierarchy Look Like in a Policy?

Figure 17 illustrates many of the topics that have been previously discussed. Note the rule hierarchy. Items that should not be stored on SharePoint are deleted first before processing the other rules. Small files would not be archived regardless of the date or if a user manually tagged a file to be archived.

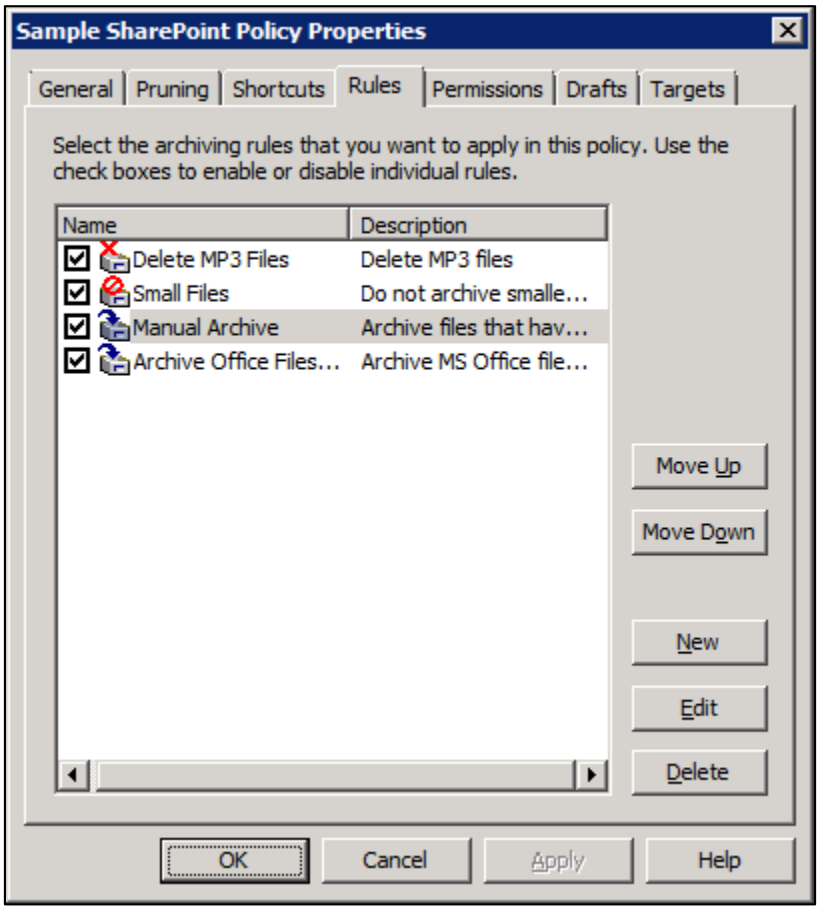

Figure 17 - Archiving rule hierarchy

Archiving for eDiscovery

A copy of all files in SharePoint can be preserved in Enterprise Vault for compliance and discovery purposes. Keep in mind that it is not necessary to create shortcuts at the time of archive. Original items can be replaced with shortcuts at a later time. For eDiscovery purposes, an "archive everything" policy is a valid use case when needing to capture all files.

Using Enterprise Vault for SharePoint to Accelerate Migrations

Enterprise Vault can be leveraged for both in-place and database-attached migration scenarios. Archiving content with Enterprise Vault before the migration will accelerate the process as the content database will be smaller.

Specific steps when upgrading from SharePoint 2007 to 2010:

- Upgrade the Enterprise Vault environment to at least Enterprise Vault 9.0 as previous versions will not support SharePoint 2010
- Replace content with shortcuts where desired
- Upgrade SharePoint from 2007 to 2010
- Re-install the Enterprise Vault SharePoint components after the upgrade to SharePoint 2010 has taken place
- If any SharePoint web applications were renamed as part of the migration, the old SharePoint targets in Enterprise Vault will need to be deleted and re-added with the new web application target names

About Symantec:

Symantec is a global leader in providing storage, security and systems management solutions to help consumers and organizations secure and manage their information-driven world.

Our software and services protect against more risks at more points, more completely and efficiently, enabling confidence wherever information is used or stored.

For specific country offices and contact numbers, please visit our Web site: **www.symantec.com**

Symantec Corporation
World Headquarters
350 Ellis Street
Mountain View, CA 94043 USA
+1 (650) 527 8000
+1 (800) 721 3934

Copyright © 2012 Symantec Corporation. All rights reserved. Symantec and the Symantec logo are trademarks or registered trademarks of Symantec Corporation or its affiliates in the U.S. and other countries. Other names may be trademarks of their respective owners.